UNDERSTANDING ANALYSIS OF COVARIANCE

Shayle R. Searle

Biometrics Unit, Cornell University, Ithaca, N. Y., U. S. A.

BU-1273-M* March 1995

ABSTRACT

The concepts of analysis of covariance are reviewed by emphasizing that it is just a combination of analysis of variance and regression. Writing the familiar linear model equation for data vector $\mathbf{y}$ as $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ where $\mathbf{X}$ is an incidence matrix (of elements 0 and 1) and $\boldsymbol{\beta}$ is a vector of main effects and interactions, the analysis of covariance is simply described as an extension of $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ to $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zb}$. Each column of $\mathbf{Z}$ is a vector, corresponding to $\mathbf{y}$, of observations on a regressor variable, namely a covariate, and $\mathbf{b}$ is a vector of coefficients, one for each covariate. Those coefficients are often called slopes (of the regression variables).

Based on this general formulation, computing formulae for estimation, sampling variances, tests of hypotheses, and some special cases are shown.

---

1. INTRODUCTION

Understanding analysis of covariance has often been found difficult in the learning and use of statistical methods. This is largely the result of being taught concepts in close combination with the teaching of the calculations required for analysis of covariance. Moreover, both concepts and calculations often come hurriedly at the very end of courses on statistical methods, or design of experiments, or introductory analysis of variance; and as a result, analysis of covariance gets squeezed and becomes poorly assimilated and understood.

Keywords and phrases: *Analysis of covariance, Analysis of variance, Factors, Regression, Estimating, Hypothesis testing.*

* Paper invited for opening the special session on "Analysis of Covariance" at the 41st Annual Meeting of the Germany Region of the International Biometric Society, held at Hohenheim University, 14 March, 1995.

So the purpose of this review is to focus on understanding the concepts of analysis of covariance: for they are not difficult. And from this understanding it is easy to go to general formulae for calculating analyses of variance for fitting models with covariates, and for calculating estimates of parameters and tests of hypotheses. These general formulae can then readily be used to derive results for many special cases.

Although this understanding of analysis of covariance relies upon familiarity with the matrix-vector formulation of the familiar linear model $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, which maybe not everyone enjoys, the concepts thus described and the general formulae obtained do apply very generally, and are widely applicable. That, and the ready availability of computers for doing the calculations, means that we do not necessarily have to worry about specific calculation formulae for special cases. Furthermore, calculations that were too complicated in pre-computer days are now easily achieved, and we do not have to worry about details of that horrible arithmetic. In addition, those details do not have to crowd out our understanding of concepts. And finally, because the general formulation and resulting formulae are so general, they can be applied to specific new models for which the arithmetic used to be hopelessly impractical.

THE 1-WAY CLASSIFICATION

2.1. Notation and examples

As a beginning, consider data that are grouped into classes, such as occupations, or diets or varieties. For what we call the response variable, y_{ij} will be the j'th observation on that variable in class i. And, corresponding to y_{ij}, will be observation z_{ij} on a covariate. (Although the letter x is often used for covariables, we here use z, in order to save X for a traditional matrix usage.) In this context, "class i" will be the i'th level of whatever factor is being considered as influencing the magnitude of y_{ij}. Thus in the examples of Table 1, class i is, respectively, the i'th occupation, the i'th calf diet, and the i'th variety of wheat. The first example concerns the effect of occupation on obesity

TABLE 1. Three examples of a 1-way classification, each with one covariate.

Factor (class)	Data from j'th observational unit in the i'th level of the factor			
4 occupations (i = 1, 2, 3 or 4)	Response Covariate	y_{ij}: z_{ij}:	obesity age	of worker j in occupation i
5 calf diets (i = 1, 2, 3, 4 or 5)	Response Covariate	y_{ij}: z_{ij}:	after one week initial	weight of calf j in diet i
6 varieties of corn (i = 2 or 3)	Response Covariate	y_{ij}: z_{ij}:	yield of corn number of plants	in plot j of variety i

Number of classes: a $\qquad i = 1, 2, \cdots, a$

Number of observations in class i: n_i $\qquad j = 1, 2, \cdots, n_i$.

(the response variable), the covariate being age, with data coming from a number of workers in each occupation. The second example deals with the effect of diet on calf weight after calves have been on a diet for one week, with initial weight as a covariate, and data from several calves on each diet. The third example concerns varieties of corn varying in yield (per plot), where the covariate is number of plants per plot. In all cases, and in general, for the 1-way classification we take

$$y_{ij} = \text{observed response on j'th observational unit in class i}$$

$$z_{ij} = \text{observed covariate on j'th observational unit in class i}$$

for a classes and n_i data pairs (y_{ij}, z_{ij}) in class i so that

$$i = 1, 2, \cdots, a \quad \text{and} \quad j = 1, 2, \cdots, n_i \ . \tag{1}$$

Then for totals and averages we use

$$y_{i\cdot} = \sum_{j=1}^{n_i} y_{ij} \quad \text{and} \quad \bar{y}_{i\cdot} = y_{i\cdot}/n_i \quad ;$$

and

$$y_{\cdot\cdot} = \sum_{i=1}^{a} \sum_{j=1}^{n_i} y_{ij} = \sum_{i=1}^{a} n_i \bar{y}_{i\cdot} \quad \text{and} \quad \bar{y}_{\cdot\cdot} = y_{\cdot\cdot} \Big/ \sum_{i=1}^{a} n_i \ .$$

Similar notation is used for the covariate data, the z_{ij}.

2.2. Analysis of variance of response variable

To lead up to our general description of analysis of covariance, we look first at how it is often presented for the simplest case possible, the 1-way classification with balanced data; i.e., the same number of observations, n, in each class, meaning $n_i = n$ for all $i = 1, 2, \cdots, a$, and $j = 1, 2, \cdots, n$ for every class. Then the familiar sums of squares for the analysis of variance of the response variable data are those shown in Table 2.

TABLE 2. Sums of squares for a 1-way classification with balanced data.

Source	d.f.	Sum of Squares
Classes	$a - 1$	$B_{yy} = \sum_{i=1}^{a} n(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$
Residual	$a(n - 1)$	$W_{yy} = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{\cdot\cdot})^2$
Total	$an - 1$	$T_{yy} = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{\cdot\cdot})^2$

On assuming that the response variable is normally distributed, the customary F-statistic calculated from the Table 1 sums of squares, namely

$$F = \frac{B_{yy}/(a - 1)}{W_{yy}/a(n - 1)} , \tag{2}$$

provides a test of the hypothesis that the population class means are equal.

2.3. Traditional analysis of covariance

Now extend the sums of squares of the y_{ij}-values in Table 2 to sums of squares of the z_{ij}-values and to sums of products of the y_{ij} and z_{ij}-values, summarized in Table 3.

TABLE 3. Sums of squares and products used as the basis for analysis of covariance of a 1-way classification with balanced data.

Source	d.f.	Sum of Squares and Products		
		y^2	z^2	yz
Classes	a − 1	B_{yy}	B_{zz}	B_{yz}
Residual	a(n − 1	W_{yy}	W_{zz}	W_{yz}
Classes	a − 1	T_{yy}	T_{zz}	T_{yz}

The traditional presentation of analysis of covariance then involves making adjustments to W_{yy} and T_{yy} to get W'_{yy} and T'_{yy} as shown in Table 4, together with $B'_{yy} = T'_{yy} - W'_{yy}$.

TABLE 4. Sums of squares for the analysis of covariance for a 1-way classification with balanced data.

Source	d.f.	Sum of Squares
Classes	a − 1	$B'_{yy} = T'_{yy} - W'_{yy}$
Residual	a(n − 1) − 1	$W'_{yy} = W_{yy} - W^2_{yz}/W_{zz}$
Total	an − 2	$T'_{yy} = T_{yy} - T^2_{yz}/T_{zz}$

The conclusion usually drawn from Table 4 is that

$$F' = \frac{B'_{yy}/(a-1)}{W'_{yy}/[a(n-1)-1]} \tag{3}$$

does, under normality, provide a test of the hypothesis that class means, adjusted for the covariate, are equal.

2.4. Questions

This description of the analysis of covariance is similar to that used in many places: e.g., Snedecor and Cochran (1989, Section 18.2) and Winer (1981, Section 10.2). Yet I find it has many deficiencies. True, the sum of squares T'_{yy} labeled 'Total' in Table 4 is simply a residual sum of squares after fitting a linear regression on z, namely

$$T'_{yy} = T_{yy} - \frac{T_{yz}^2}{T_{zz}} = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 - \frac{\left[\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})(z_{ij} - \bar{z}_{..})^2\right]^2}{\sum_{i=1}^{a}\sum_{j=1}^{n}(z_{ij} - \bar{z}_{..})^2} \;,$$

so it is, in some sense, adjusting for the covariate. And in somewhat the same way W'_{yy} represents a similar adjustment. Other than this, though, there seems to be no directly clear reason why F′ of (3) is the analog of (2) for testing that class means adjusted for the covariate are equal. Indeed, a number of questions arise from Table 4 and F′ of (3) that do not seen to be directly answered by considering Table 4 as just a natural extension of Table 2, the usual analysis of variance of the y_{ij}s. These questions include the following.

I. How do we recognize F′ as having an F-distribution?

II. How do we *know* that F′ tests the hypothesis that class means adjusted for the covariate are equal?

III. Just exactly what is a "class mean adjusted for the covariate"?

IV. How is the effect of the covariate estimated, both in Table 4 (for balanced data), and for unbalanced data?

V. How is Table 4 to be adapted for unbalanced data?

VI. How is Table 4 to be adapted to handling the covariate differently for each class?

VII. How is the methodology of Table 4 adapted to data more complicated than those of a completely randomized experiment (a 1-way classification); e.g., to a randomized complete block experiment (a 2-way classification)?

VIII. How does that methodology extend to having more than one covariate?

IX. What is the analysis for a combination of V, VI, VII and VIII?

These are all important questions. Answers to them are not obvious from Table 4. And yet the answers are computable with today's fast and efficient computing facilities. This means that there are ways of formulating analysis of covariance so that the answers can be expressed algebraically (and hence can be computed). We therefore need to understand the basis of this, and in doing so we need not worry about the arithmetic details because computer programs will take care of those. And fortunately, so long as one understands a little matrix algebra, describing and understanding the basis of analysis of covariance is not difficult.

3. COMBINING FACTORS and REGRESSION

3.1. The basic model equation

The essential idea of analysis of covariance is that it is a combination of factors and regression. Begin by writing the expected value of y_{ij} of the 1-way classification of the preceding section as

$$E(y_{ij}) = \mu + \alpha_i \, . \tag{4}$$

Then on defining

$$e_{ij} = y - E(y_{ij})$$

we have

$$y_{ij} = \mu + \alpha_i + e_{ij} \, , \tag{5}$$

the familiar model equation for the 1-way classification.

The extension of (4) to the complete data arrayed as a vector is

$$E(\mathbf{y}) = \mathbf{X}\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_a \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} \quad \text{for} \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_a \end{bmatrix} . \tag{6}$$

This equation, $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, on defining $\mathbf{e} = \mathbf{y} - E(\mathbf{y})$ as a vector of random error terms, leads at once to the familiar model equation for linear models:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \, . \tag{7}$$

This form of equation is, as is well-known, generally applicable to all traditional analysis of variance linear models where its $\mathbf{X}$ is a matrix of zeros and ones and its $\boldsymbol{\beta}$ is a vector of parameters, of which (6) is an example. Those parameters, as in (6), are usually a general mean μ, and effects for the levels of each main effects factor and each interaction that occurs in whatever model is being used. For example, for data from a randomized complete block experiment we might take $E(y_{ijk}) = \mu + \tau_i + \theta_j + \varphi_{ij} + e_{ijk}$, whereupon $\boldsymbol{\beta}$ of (7) would consist of μ, and the τs, θs and φs, representing a mean and main effects τ_i and θ_j and interactions φ_{ij}.

But $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ can also be used for regression where $\boldsymbol{\beta}$ has μ and regression coefficients as its elements. To distinguish this $\mathbf{X}\boldsymbol{\beta}$ from its use for main effects and interactions of the previous paragraph, let us use $\mathbf{Zb}$ for the covariates that will occur in analysis of covariance. Then $\mathbf{b}$ will be the vector of regression coefficients (but not μ) and $\mathbf{Z}$ will have columns that are observed covariates – or regressor variables if one prefers that name. Then analysis of covariance combines $\mathbf{X}\boldsymbol{\beta}$ for main effects and interactions with $\mathbf{Z}\boldsymbol{\beta}$ for covariates to have a model equation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zb} + \mathbf{e} \ . \tag{8}$$

For example, in the 1-way classification with a single covariate, (5) would become

$$\mathbf{y}_{ij} = \mu + \alpha_i + b_1 z_{ij} + \mathbf{e}_{ij} \ . \tag{9}$$

giving (8) as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z}b_1 + \mathbf{e} \ . \tag{10}$$

The $\mathbf{X}\boldsymbol{\beta}$ here is as in (6), and $\mathbf{zb}_1$ is the $\mathbf{Zb}$ of (8) with $\mathbf{Z}$ being just the single column vector $\mathbf{z}$ of (10), containing the z_{ij}-values of (9); and $\mathbf{b}$ of (8) would be just the scalar b_1 of (9).

Were there to be a second covariate, w_{ij} say, (9) would be of the form

$$\mathbf{y}_{ij} = \mu + \alpha_i + b_1 z_{ij} + b_2 w_{ij} + \mathbf{e}_{ij}$$

giving

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + [\mathbf{z} \quad \mathbf{w}]\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \mathbf{e}$$

rewritten as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zb} + \mathbf{e}$ of (8). This model equation is a simple basis for understanding analysis of covariance: a linear model having two parts, $\mathbf{X}\boldsymbol{\beta}$ for main effect and interaction factors, and $\mathbf{Zb}$ for covariates (or regression variables).

3.2. Accounting for a covariate mean

Readers' reaction to the model equation (9) may well be that where there is z_{ij} there should be $(z_{ij} - \bar{z}_{..})$ so that (9) would then be

$$y_{ij} = \mu + \alpha_i + b_1(z_{ij} - z_{..}) + e_{ij} \ . \tag{11}$$

It is certainly true that (11) is to be found in many places [e.g., Snedecor and Cochran (1989) and Winer (1981)] as the appropriate model equation. Neither (9) nor (11) is wrong. They are equivalent, and either can be used. I prefer (9) because it is simpler, and also because it strikes me as distasteful to have a sample (observed) mean as part of a model equation.

3.3. Flexibility of the general model equation

Flexibility of the model equation (8) will be illustrated using the small numerical example of Table 6.1 in Searle (1987), shown here as Table 5.

TABLE 5. Data for illustrating some special cases of $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zb} + \mathbf{e}$, for the 1-way classification.

Class	Response variable y_{ij}	Covariate 1 z_{ij}	Covariate 2 w_{ij}
i = 1	74	3	22
	68	4	26
	77	5	21
i = 2	76	2	21
	80	4	27
i = 3	87	3	20
	91	7	24

To begin, we illustrate $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ of (7) for just the analysis of variance of the y_{ij} data of Table 5. This is

$$\mathbf{y} = \begin{bmatrix} 74 \\ 68 \\ 77 \\ 76 \\ 80 \\ 87 \\ 91 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \mathbf{e} \ ,$$

where a dot in a matrix represents zero. Now, for the same data we show model equations of the form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e}$ for six different analysis of covariance models.

(i) <u>Single covariate, traditional treatment</u> $\quad y_{ij} = \mu + \alpha_i + b_1 z_{ij} + e_{ij}$.

$$\mathbf{y} = \begin{bmatrix} 74 \\ 68 \\ 77 \\ 76 \\ 80 \\ 87 \\ 91 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + b_1 \begin{bmatrix} 3 \\ 4 \\ 5 \\ 2 \\ 4 \\ 3 \\ 7 \end{bmatrix} + \mathbf{e}\,.$$

This provides an answer to question V asked earlier, about dealing with unbalanced data.

(ii) <u>Two covariates, traditional treatment</u> $\quad y_{ij} = \mu + \alpha_i + b_1 z_{ij} + b_2 w_{ij}$.

$$\mathbf{y} = \begin{bmatrix} 74 \\ 68 \\ 77 \\ 76 \\ 80 \\ 87 \\ 91 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} 3 & 22 \\ 4 & 26 \\ 5 & 21 \\ 2 & 21 \\ 4 & 27 \\ 3 & 20 \\ 7 & 24 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \mathbf{e}\,.$$

There are two covariates, two bs, and two columns of **Z**, each of which contains, corresponding to the y_{ij}, the values of a covariate. This leads to answering question VIII asked earlier.

(iii) <u>Single covariate, each class having its own covariate coefficient</u> $\quad y_{ij} = \mu + \alpha_i + b_i z_{ij} + e_{ij}$.

$$\mathbf{y} = \begin{bmatrix} 74 \\ 68 \\ 77 \\ 76 \\ 80 \\ 87 \\ 91 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} 3 & \cdot & \cdot \\ 4 & \cdot & \cdot \\ 5 & \cdot & \\ \cdot & 2 & \cdot \\ \cdot & 4 & \cdot \\ \cdot & \cdot & 3 \\ \cdot & \cdot & 7 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + \mathbf{e}\,.$$

Here there are three columns in **Z**, because there are three classes in the 1-way classification and there is to be a different b (a different "slope") for each class. The zero elements in **Z** are no cause for concern whatever. When later we discuss properties of columns of **Z** they apply equally as well to those containing zeros as to those not containing zeros.

(iv) A variation on (iii).

Suppose in (iii) we wished to test the hypothesis $H : b_2 = b_3$. The easiest way to calculate an F-statistic for this purpose would be to use the standard formula such as in Searle [1971, p. 190, equation (70)] and Searle [1987, p. 291, equation (146) – using (144)]. An alternative would be to use the difference between two sums of squares, one for fitting the model of (iii) and one for fitting

$$\mathbf{y} = \begin{bmatrix} 74 \\ 68 \\ 77 \\ 76 \\ 80 \\ 87 \\ 91 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} 3 & \cdot \\ 4 & \cdot \\ 5 & \cdot \\ \cdot & 2 \\ \cdot & 4 \\ \cdot & 3 \\ \cdot & 7 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \mathbf{e} \ .$$

Both this model equation and that in (iii) produce answers to the earlier-asked question VI.

(v) Two covariables, treated differently

An illustrative answer to question IX is to show the model equation for covariate z_{ij} having a different slope for each class, as in (iii), and covariate w_{ij} having the same slope, as in (ii). This has model equation $y_{ij} = \mu + \alpha_i + b_i z_{ij} + b_4 w_{ij} + e_{ij}$.

$$\mathbf{y} = \begin{bmatrix} 74 \\ 68 \\ 77 \\ 76 \\ 80 \\ 87 \\ 91 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} 3 & \cdot & \cdot & 22 \\ 4 & \cdot & \cdot & 26 \\ 5 & \cdot & \cdot & 21 \\ \cdot & 2 & \cdot & 21 \\ \cdot & 4 & \cdot & 27 \\ \cdot & \cdot & 3 & 20 \\ \cdot & \cdot & 7 & 24 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} + \mathbf{e} \ .$$

This is, of course, ridiculous for this particular data set because there are only seven data values and seven estimable parameter functions. But it does illustrate the flexibility of the model equation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e}$.

(vi) A 2-way classification, with one covariate

Suppose the data of Table 5 are rearranged as unbalanced data from a 3×3 2-way classification as shown in Table 7.

TABLE 6. Data of Table 5 as coming from a 3 × 3 layout.

	y_{ij}			z_{ij}		
	j = 1	j = 2	j = 3	j = 1	j = 2	j = 3
i = 1	74	68	77	3	4	5
i = 2		76	80		2	4
i = 3	87	91		3	7	

For fitting a no-interaction model for factor effects (rows and columns in Table 7) and a different slope for z_{ij} for each column, the model equation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e}$ would be

$$y_{ij} = \mu + \alpha_i + \beta_j + b_j z_{ij} + e_{ij} \ .$$

$$\mathbf{y} = \begin{bmatrix} 74 \\ 68 \\ 77 \\ 76 \\ 80 \\ 87 \\ 91 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdot & \cdot & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & 1 & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 \\ 1 & \cdot & 1 & \cdot & \cdot & 1 & \cdot \\ 1 & \cdot & 1 & \cdot & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & 1 & 1 & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot & 1 & \cdot \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} 3 & \cdot & \cdot \\ \cdot & 4 & \cdot \\ \cdot & \cdot & 5 \\ \cdot & 2 & \cdot \\ \cdot & \cdot & 4 \\ 3 & \cdot & \cdot \\ \cdot & 7 & \cdot \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + \mathbf{e} \ .$$

This is a grossly over-parameterized model for the seven data values: but again, it is highly illustrative of the flexibility of the model equation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e}$ being able to accommodate a wide variety of analysis of covariance models.

These illustrations also demonstrate that answers to questions V through IX can be provided: V is illustrated by all of (i) through (vi), VI by (iii) and (iv), VII by (vi), VIII by (ii) and IX by (v). We therefore proceed to a summary of some of the calculations.

4. A SUMMARY OF CALCULATION FORMULAE

The impressive aspect of the preceding illustrations of the model equation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zb} + \mathbf{e}$ is that it is available for a very wide variety of applications – indeed all linear models involving factors and covariables. And in saying this, recall that for linear models "linear" means linear in the parameters, so that one can even deal with polynomial functions of covariates, including such uncustomary terms such as powers and products of covariates; e.g., z_{ij}^2 and $z_{ij}w_{ij}$. To calculate estimates, their sampling variances, F-ratios, tests of hypotheses and other useful statistics, we therefore need only the one starting point:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zb} + \mathbf{e} \quad . \tag{12}$$

4.1. Estimation

Recall for a moment that in the commonplace linear model where $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, least squares estimation starts with the normal equations $\mathbf{X'X}\boldsymbol{\beta}^o = \mathbf{X'y}$ having a solution $\boldsymbol{\beta}^o = (\mathbf{X'X})^-\mathbf{X'y}$ for $(\mathbf{X'X})^-$ being any generalized inverse of $\mathbf{X'X}$ satisfying $\mathbf{X'X(X'X)^-X'X} = \mathbf{X'X}$. From these, all manner of useful results emanate.

We use those results for analysis of covariance by using (12) in place of $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$. To do so, first rewrite (12) as

$$\mathbf{y} = [\mathbf{X} \quad \mathbf{Z}]\begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix} + \mathbf{e} \ .$$

The resulting normal equations are

$$\begin{bmatrix} \mathbf{X'} \\ \mathbf{Z'} \end{bmatrix}[\mathbf{X} \quad \mathbf{Z}]\begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'} \\ \mathbf{Z'} \end{bmatrix}\mathbf{y} \ ,$$

more revealingly expressed as

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} \end{bmatrix}\begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix} . \tag{13}$$

A useful solution of (13) comes from making two very practical, seldomly-violated, assumptions about the columns of $\mathbf{Z}$: that they are linearly independent of each other and of columns of $\mathbf{X}$. Then a solution of (13) is for

$$\mathbf{M} = \mathbf{I} - \mathbf{X(X'X)^-X} = \mathbf{M'} = \mathbf{M}^2 \quad \text{with} \quad \mathbf{MX} = \mathbf{0}$$

and

$$\mathbf{R} = \mathbf{MZ}$$

that

$$\hat{\mathbf{b}} = (\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'\mathbf{y} = (\mathbf{R}'\mathbf{R})^{-1}(\mathbf{y} - \hat{\mathbf{y}}) \quad \text{for} \quad \hat{\mathbf{y}} = \mathbf{M}\mathbf{y}$$

and

$$\tilde{\beta} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'(\mathbf{y} - \mathbf{Z}\hat{\mathbf{b}}) = \beta^{o} - (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Z}\hat{\mathbf{b}} \ . \tag{14}$$

The distinguishing notation of $\hat{\mathbf{b}}$ (with a hat) and $\tilde{\beta}$ (with a tilde) is as follows. For given data and model (i.e., given **y**, **X** and **Z**), $\hat{\mathbf{b}}$ is unique and is the best linear unbiased estimator (BLUE) of **b**. In contrast, $\tilde{\beta}$ is not invariant to the choice of $(\mathbf{X}'\mathbf{X})^{-}$; but $\mathbf{X}\tilde{\beta}$ is invariant and

$$\text{BLUE}(\mathbf{X}\beta) = \mathbf{X}\tilde{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'(\mathbf{y} - \mathbf{Z}\hat{\mathbf{b}}) \ . \tag{15}$$

4.2. Using residuals

In the no-covariate linear model based on $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$, the predicted value of **y** is

$$\hat{\mathbf{y}} = \mathbf{X}\beta^{o} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} \ .$$

Then the residual vector is

$$\mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \hat{\mathbf{y}} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}']\mathbf{y} = \mathbf{M}\mathbf{y} \ . \tag{16}$$

Now, in the with-covariate model, consider **R** of

$$\hat{\mathbf{b}} = (\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'\mathbf{y} \qquad \text{where} \qquad \mathbf{R} = \mathbf{M}\mathbf{Z} \ . \tag{17}$$

For $\mathbf{z}_t$ being the t'th column of **Z**, the t'th column of **R** is $\mathbf{M}\mathbf{z}_t$. By analogy with (16) we see that $\mathbf{M}\mathbf{z}_t$ is $\mathbf{r}_t = \mathbf{z}_t - \hat{\mathbf{z}}_t$ where $\hat{\mathbf{z}}_t$ is the predicted value of $\mathbf{z}_t$ after (for computing purposes only) fitting $\mathbf{z}_t$ to the no-covariate model $\mathrm{E}(\mathbf{z}) = \mathbf{X}\beta$. This is the mnemonic nature of the symbol **R** used for $\mathbf{M}\mathbf{Z}_t$ in $\hat{\mathbf{b}}$. Each column of **R** is the vector of residuals after fitting the corresponding column of **Z** to the no-covariate model; i.e., after fitting $\mathbf{z}_t = \mathbf{X}\beta + \mathbf{e}$. It is strongly emphasized that this fitting is only a computational procedure. But it applies quite generally, even when columns of **Z** contain zeros, as for example in illustration (iii). Moreover, by the two assumptions made earlier about linear independence of columns of **Z**, we are always assured that $(\mathbf{R}'\mathbf{R})^{-1}$ exists.

4.3. Analysis of variance tables

Analysis of covariance can be used in two different ways, for distinctly different purposes. (a) For economists and others concerned in trends, but where factors also occur, interest often lies in testing for trends (i.e., for **b**), adjusted for factors. (b) But for many experimental situations, where factors are the important features and covariates are of secondary importance, interest will lie in factors, adjusted for covariates. We show an analysis of variance table for each of these situations. To do so, it is necessary to introduce further notation.

$\mathbf{X}^+$ being the Moore-Penrose of $\mathbf{X}$ gives $\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}' = \mathbf{X}\mathbf{X}^+$.

N = number of data values in $\mathbf{y}$.

$\bar{y}$ = mean of the N data values in y.

$\mathbf{C} = \mathbf{I}_N - \bar{\mathbf{J}}_N$ where $\mathbf{I}_N$ is an identity matrix of order N, and $\bar{\mathbf{J}}_N$ is $N \times N$ with all elements $1/N$.

$\mathbf{X} = [\mathbf{1}_N \quad \mathbf{X}_1]$, so that $\mathbf{X}_1$ is $\mathbf{X}$ without its first column.

r_X = rank of $\mathbf{X}$ = rank of $\mathbf{X}_1$.

$\mathbf{1}_N$ is $N \times 1$ with every element unity.

$\boldsymbol{\beta}_1$ is every element of $\boldsymbol{\beta}$ except μ, so that $\boldsymbol{\beta} = \begin{bmatrix} \mu \\ \boldsymbol{\beta}_1 \end{bmatrix}$.

$\mathbf{y}'\mathbf{y}$ = sum of squares of each observation in $\mathbf{y}$.

$\mathfrak{Z}$ = $\mathbf{Z}$ with each column $\mathbf{z}_t$ replaced by $\mathbf{z}_t - \bar{z}_t\mathbf{1}_N$.

c = number of covariates, columns in $\mathbf{Z}$, and elements in $\boldsymbol{\beta}$.

Each associated hypothesis in Table 7 is the hypothesis tested, under normality, by the F-statistic calculated as the ratio of the corresponding mean square divided by the residual mean square. For example, in the second line of the table,

$$F = \frac{\mathbf{y}'\mathbf{R}\mathbf{R}^+\mathbf{y}/c}{(\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{X}^+\mathbf{y} - \mathbf{y}'\mathbf{R}\mathbf{R}^+\mathbf{y})/(N - r_X - c)} \qquad \text{tests} \qquad H : \mathbf{b} = \mathbf{0} \ .$$

Clearly, only the second lines of parts (a) and (b) of the table test clean hypotheses, i.e., hypotheses of practical use.

TABLE 7. Analyses of variance for the model equation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{B} + \mathbf{e}\ .$$

Source of Variation	d.f.	Sum of Squares	Associated Hypothesis
		(a) Fitting factors and then covariates	
Factors, adjusted for mean	$r_X - 1$	$\mathbf{y}'\mathbf{X}\mathbf{X}^{+}\mathbf{y} - N\bar{y}^2$	$H: \mathbf{C}\mathbf{X}_1[\boldsymbol{\beta}_1 + (\mathbf{X}_1'\mathbf{C}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{C}\mathbf{Z}\mathbf{b}] = \mathbf{0}$
Covariates, adjusted for mean and factors	c	$\mathbf{y}'\mathbf{R}\mathbf{R}^{+}\mathbf{y}$	$H: \mathbf{b} = 0$
Residual	$N - r_X - c$	$\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{X}^{+}\mathbf{y} - \mathbf{y}'\mathbf{R}\mathbf{R}^{+}\mathbf{y}$	
Total	$N - 1$	$\mathbf{y}'\mathbf{y} - N\bar{y}^2$	
		(b) Fitting covariates and then factors	
Covariates and mean, adjusted for mean	c	$\mathbf{y}'\mathfrak{Z}\mathfrak{Z}^{+}\mathbf{y}$	$H: (\mathfrak{Z}'\mathfrak{Z})^{-1}\mathfrak{Z}'\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{b} = \mathbf{0}$
Factors adjusted for mean and covariates	$r_X - 1$	$\mathbf{y}'\mathbf{X}\mathbf{X}^{+}\mathbf{y} + \mathbf{y}'\mathbf{R}\mathbf{R}^{+}\mathbf{y} - \mathbf{y}'\mathfrak{Z}\mathfrak{Z}^{+}\mathbf{y} - N\bar{y}^2$	$H:$ all elements of $\mathbf{X}_1\boldsymbol{\beta}_1$ equal
Residual	$N - r_X - c$	$\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{X}^{+}\mathbf{y} - \mathbf{y}'\mathbf{R}\mathbf{R}^{+}\mathbf{y}$	
Total	$N - 1$	$\mathbf{y}'\mathbf{y} - N\bar{y}^2$	

4.4 Sampling variances

On taking the variance-covariance matrix of $\mathbf{e}$ as $\text{var}(\mathbf{e}) = \sigma^2\mathbf{I}$, the variance and covariance properties of $\hat{\mathbf{b}}$ and $\tilde{\boldsymbol{\beta}}$ are as follows.

$$\text{var}(\hat{\mathbf{b}}) = (\mathbf{R}'\mathbf{R})^{-1}\sigma^2 \quad \text{and} \quad \text{cov}(\mathbf{y},\ \hat{\mathbf{b}}) = \mathbf{R}(\mathbf{R}'\mathbf{R})^{-1}\sigma^2\ ;$$

and

$$\text{var}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'[\mathbf{I} + \mathbf{Z}(\mathbf{R}'\mathbf{R})^{-1}\mathbf{Z}']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\sigma^2\ ,$$

after correcting a minus sign to a plus in Searle [1987, p. 423, equation (25)]. Other covariances that may be needed are easily derived from these expressions.

And from Table 7 we estimate σ^2 by

$$\hat{\sigma}^2 = \frac{\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{X}^{+}\mathbf{y} - \mathbf{y}'\mathbf{R}\mathbf{R}^{+}\mathbf{y}}{\text{N} - \text{r}_{\boldsymbol{X}} - \text{c}}\ .$$

4.5. Three general hypotheses

In addition to the hypotheses shown in Table 7, three other general hypotheses can be used to yield hypotheses of particular interest. Each of these hypotheses is tested by the F-statistic

$$\text{F} = \frac{\text{Q}}{\text{r}_{\boldsymbol{K}}\hat{\sigma}^2}\ ,$$

where Q is a sum of squares and $\text{r}_{\boldsymbol{K}}$ is the rank of the full row rank matrix $\mathbf{K}'$ that is part of the hypothesis statement. The three hypotheses are as follows.

(A) $\text{H} : \mathbf{K}'\mathbf{b} = \mathbf{m}$, for any $\mathbf{K}'$, for which

$$\text{Q} = (\mathbf{K}'\hat{\mathbf{b}} - \mathbf{m})'[\mathbf{K}'(\mathbf{R}'\mathbf{R})^{-1}\mathbf{K}]^{-1}(\mathbf{K}'\hat{\mathbf{b}} - \mathbf{m})\ .$$

(B) $\text{H} : \mathbf{K}'\boldsymbol{\beta} = \mathbf{m}$, with $\mathbf{K}' = \mathbf{T}'\mathbf{X}$ for some $\mathbf{T}'$, for which

$$\text{Q} = (\mathbf{K}'\tilde{\boldsymbol{\beta}} - \mathbf{m})'\left\{\mathbf{K}'\mathbf{X}^{+}[\mathbf{I} + \mathbf{Z}(\mathbf{R}'\mathbf{R})^{-1}\mathbf{Z}']\mathbf{X}^{+\prime}\mathbf{K}\right\}^{-1}(\mathbf{K}'\tilde{\boldsymbol{\beta}} - \mathbf{m})\ .$$

(C) $\text{H} : \mathbf{K}'[\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}\mathbf{Z}\mathbf{b}] = \mathbf{0}$ with $\mathbf{K}' = \mathbf{T}'\mathbf{X}$ for some $\mathbf{T}'$, for which

$$\text{Q} = \boldsymbol{\beta}^{o\prime}\mathbf{K}[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{K}]^{-1}\mathbf{K}'\boldsymbol{\beta}^{o}\ .$$

Cases (A) and (B), being hypotheses about elements of $\mathbf{b}$ and $\boldsymbol{\beta}$, respectively, are of particular interest.

5. SPECIAL CASES

There is, of course, an endless number of special cases of the general results in the preceding section. Explicit formulae for the following thirteen cases are given in Searle (1987, Sections 11.5 – 11.7).

(a) 1-way classification, 2 covariates

– i. Single slope for each covariate

$$E(y_{ij}) = \mu + \alpha_i + bz_{ij} + b^*w_{ij}$$

– ii. Intra-class slopes for one covariate, single slope for the other

$$E(y_{ij}) = \mu + \alpha_i + b_iz_{ij} + b^*w_{ij}$$

–iii. Intra-class slopes for each covariate

$$E(y_{ij}) = \mu + \alpha_i + b_iz_{ij} + b_i^*w_{ij}$$

(b) 2-way classification, 1 covariate, single slope

– i. With interaction: $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$

$$E(y_{ijk}) = \mu_{ij} + \lambda z_{ijk}$$

– ii. No interaction: $\mu_i = \mu + \alpha_i$

$$E(y_{ijk}) = \mu_i + \tau_j + \lambda z_{ijk}$$

(c) 2-way classification (rows and columns), 2 covariates, multiple slopes

[I] Interaction models ($\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$), unbalanced data

– i. Intra-row slopes for 1 covariate

$$E(y_{ijk}) = \mu_{ij} + \lambda_iz_{ijk}$$

– ii. Intra-row slopes for 1 covariate, intra-column for another

$$E(y_{ijk}) = \mu_{ij} + \lambda_iz_{ijk} + \lambda_j^*w_{ijk}$$

–iii. Intra-row slopes for each of 2 covariates

$$E(y_{ijk}) = \mu_{ij} + \lambda_iz_{ijk} + \lambda_i^*w_{ijk}$$

– iv. Intra-row plus intra-column slopes for 1 covariate

$$E(y_{ijk}) = \mu_{ij} + (\lambda_i + \lambda_j^*)z_{ijk}$$

[II] No interactions, one observation in every cell .

The four cases of [I], but with $\mu_{ij} = \mu_i + \tau_j$, and y_{ij}, z_{ij} and w_{ij} in place of y_{ijk}, z_{ijk} and w_{ijk}.

6. THE 1-WAY CLASSIFICATION

Answers to questions V through IX of Section 2.4 are illustrated by the six examples of Section 3.3. Questions I – IV are answered here, using the 1-way classification as illustration. Its model equation, from (9), is taken as

$$y_{ij} = \mu + \alpha_i + bz_{ij} + e_{ij} \tag{18}$$

for $j = 1, 2, \cdots, n_i$ and $i = 1, 2, \cdots, a$; i.e., unbalanced data.

6.1. Analyses of variance

For (18), the notation of Table 7 is

$$\mathbf{XX}^+ = \left\{_d \, \bar{\mathbf{J}}_{n_i}\right\}, \qquad \mathbf{X}_1 = \left\{_d \, \mathbf{1}_{n_i}\right\}, \qquad \boldsymbol{\beta}_1 = \{\alpha_i\}, \qquad r_{\mathbf{X}} = a$$

$$\mathbf{Z} = \mathbf{z} = \left\{_c \, z_{ij}\right\}, \qquad \mathbf{RR}^+ = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_{..})^2 \qquad \text{and} \qquad c = 1 .$$

As a result, Table 7 simplifies to be Table 8, wherein the B, W and T symbols have the same representation as in Tables 2, 3 and 4 of Section 2 – except that they now apply to unbalanced (rather than to just balanced) data.

6.2. Test of hypotheses

Three hypotheses are worth noting. First, from part (a), which deals with fitting factors (class effects) before the covariate, two useful hypotheses can be tested using:

$$F_1 = \frac{B_{yy}/(a-1)}{W'_{yy}/(N-a-1)} \tag{19}$$

tests

$$H: \alpha_1 + b\bar{z}_{1.} = \alpha_2 + b\bar{z}_{2.} = \cdots = \alpha_i + b\bar{z}_{i.} = \cdots = \alpha_a + b\bar{z}_{a.} , \tag{20}$$

and using

$$F_2 = \frac{(W_{yz}^2/W_{zz})/1}{W'_{yy}/(N-a-1)} \tag{21}$$

tests

$$H: b = 0 . \tag{22}$$

Then, from part (b) of Table 8,

$$F_3 = \frac{B'_{yy}/(a-1)}{W'_{yy}/(N-a-1)} \tag{23}$$

tests

$$H: \alpha_1 = \alpha_2 = \cdots = \alpha_i = \cdots = \alpha_a . \tag{24}$$

TABLE 8. Analyses of variance for the model equation

$y_{ij} = \mu + \alpha_i + bz_{ij} + e_{ij}$ for $j = 1, \cdots, n_i$ and $i = 1, \cdots, a$.

Source of Variation	d.f.	Sum of Squares	Associated Hypothesis
		(a) Fitting factors and then covariates	
Factors, adjusted for mean	$a - 1$	B_{yy}	$H : \alpha_i + b\bar{z}_{i\cdot}$ equal $\forall i$
Covariates, adjusted for mean and factors	1	W_{yz}^2/W_{zz}	$H : b = 0$
Residual	$N - a - 1$	W'_{yy}	
Total	$N - 1$	T_{yy}	
		(b) Fitting covariates and then factors	
Covariates and mean, adjusted for mean	1	T_{yz}^2/T_{zz}	$H : \sum_{i=1}^{a} k_i\alpha_i + b = 0$*
Factors adjusted for mean and covariates	$a - 1$	B'_{yy}	$H : \alpha_i$ all equal
Residual	$N - a - 1$	W'_{yy}	
Total	$N - 1$	T_{yy}	

* $k_i = n_i(\bar{z}_{i\cdot} - \bar{z}_{\cdot\cdot})/T_{zz}$ with $\Sigma k_i = 0$.

It is easily seen that F_3 of (23) derived from Table 8 is identical to F′ of (3) derived from Table 4. Thus question I is answered: we know why F′ has an F-distribution. More than that, from (24) we know what the hypothesis is that is tested by $F'(= F_3)$. Thus is question II answered. Any by looking at (24) we get the impression that a "class mean adjusted for covariates" is here interpreted as α_i; this answers question III. And the F-statistic and hypothesis of (21) and (22) show how the effect of the covariate can be assessed — so answering, a little obliquely, question IV. A direct answer follows.

6.3. Estimation: and adjusted treatment means

The general estimation formula (17) for **b** reduces for the model (18) to

$$\hat{b} = W_{yz}/W_{zz} , \tag{25}$$

thus providing an estimate of the influence of the covariate z on the response variable y, with a test of $H : b = 0$ being provided in part (a) of Table 8.

And from (14) we get

$$\begin{bmatrix} \mu^0 \\ \{\alpha_i^0\} \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \{_d \ 1/n_i\} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \{_d \ \mathbf{1}'_{n_i}\} \end{bmatrix} (\mathbf{y} - \hat{b}\mathbf{z}) = \begin{bmatrix} 0 \\ \{\bar{y}_{i\cdot} - \hat{b}\bar{z}_{i\cdot}\} \end{bmatrix} ,$$

which gives, in according with standard results (e.g., Searle, 1987)

$$\text{BLUE}(\mu + \alpha_i) = \mu^0 + \alpha^0{}_i = \bar{y}_{i\cdot} - \hat{b}\bar{z}_{i\cdot} \ . \tag{26}$$

6.4. Adjusted treatment means

The phrase "adjusted treatment mean" is used by many writers in the description of the hypothesis tested by $F_3(= F')$ of (23) and (3). And from part (b) of Table 8 we see that it tests

$$H : \alpha_i \text{ all equal}$$

which is, of course, equivalent to

$$H : (\mu + \alpha_i) \text{ all equal} . \tag{27}$$

Hypotheses are always described in terms of parametric functions, and $F_3(=F')$ tests (   ) and is often described as testing equality of adjusted treatment means. This could suggest calling $\mu + \alpha_i$ of (27) an adjusted treatment mean, and so $\bar{y}_{i\cdot} - \hat{b}\bar{z}_{i\cdot}$, in being an estimator (the BLUE) of $\mu + \alpha_i$, could perhaps be called an estimated adjusted treatment mean. But this has no appeal. $\mu + \alpha_i$ of itself contains no suggestion of adjusting for the covariate, where $\bar{y}_{i\cdot} - \hat{b}\bar{z}_{i\cdot}$ very clearly does. Thus, even though F′ tests a hypothsis involving $\mu + \alpha_i$ that gets described in terms of adjusted treatment means, the $\mu + \alpha_i$ is really a population mean (ignoring the covariate), whereas its estimator $\bar{y}_{i\cdot} - \hat{b}\bar{z}_{i\cdot}$ is truly an adjusted treatment mean.

The nomenclature "adjusted treatment mean" is further clouded by the fact that some writers use it to refer to $\bar{y}_{i\cdot} - \hat{b}(\bar{z}_{i\cdot} - \bar{z}_{\cdot\cdot})$. The confusion seems destined to remain: that F′ tests a hypothesis which will continue to be described in terms of "adjusted treatment means" which is not apt for the actual form of parametric function involved in the hypothesis, but it is appropriate for its estimator.

REFERENCES

Searle, S. R. (1971) *Linear Models*, John Wiley & Sons, New York.

Searle, S. R. (1987) *Linear Models for Unbalanced Data*, John Wiley & Sons, New York.

Snedecor, G. W. and Cochran, W. G. (1989) *Statistical Methods*, 8th ed., Iowa State University Press, Ames, Iowa.

Winer, B. J. (1981) *Statistical Principles in Experimental Design*, 2nd ed., McGraw-Hill, New York.